Article 6.1

The file named Article6_1.xlsx contains data for US firms for the period 2011-2019. The variables included in the dataset are:

- Investment is the ratio of capital expenditure to total asset;
- Mtb is firm's market to book;
- Leverage is the amount of total debt to total asset;
- Cashflow is the cashflow of each firm to total asset;
- Cashtoasset is the ratio between cash and cash equivalents to total asset;
- Total Asset.

QUESTIONS:

1. The structure of the data is a panel. Sort the data by ID and time.
2. Generate a new variable "size". A good proxy for the size of the firm is the log of total asset. So, calculate "size" as the natural logarithm of total asset.
3. Some literature argues that if firms hold an amount of cash and cash equivalents not smaller than the 25% of their total assets, then they should be considered cash conservative. Construct a dummy variable, which takes the value of 1 if the value in the variable cashtoasset is equal or greater than 0.25 and 0 otherwise.
4. Calculate the summary statistics for the two new variables: size and cash_con. Comment the values obtained. Which is the probability of being cash conservative firms?
5. Derive the correlation matrix for the main variables.
6. Using the pivot table, calculate the average of investments for cash conservative and non-cash conservative firms.
7. Create a dummy variable which takes the value of 1 if size is equal or greater than its average (big size firms) and 0 otherwise (small size firms)
8. Using the pivot table, calculate the average of investments for different sizes and for cash and non-cash conservative firms.

9. Calculate the mean of the variable cashflow for each year. Plot the obtained values in an appropriate graph.

10. Sheet2 contains the variable “Sales” which is the ln of net sales, and it is employed as a proxy of firms’ revenues. Incorporate this variable into the main dataset (sheet “data”) using VLOOKUP function.

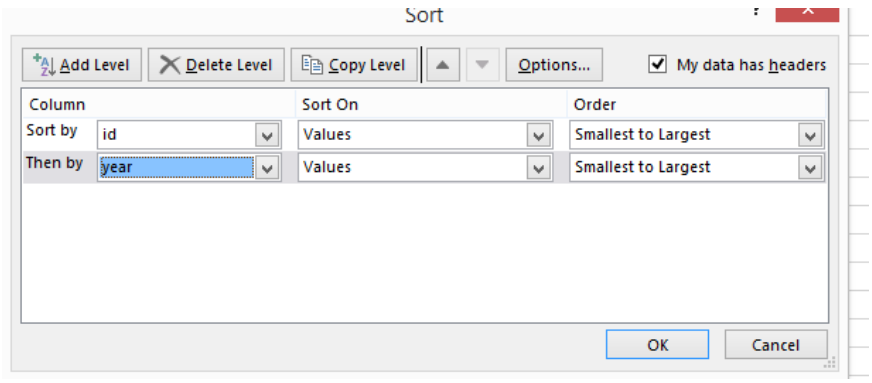
11. Estimate the following equation using OLS:

$$investment_{it} = \beta_0 + \beta_1 cashflow_{it} + \beta_2 mtb_{it} + \varepsilon_{it}$$

SOLUTIONS:

1. The structure of the data is a panel. Sort the data by ID and time.

   Solution: Select all our data -> click **sort** in Data Tab -> **sort by ID** -> **Add Level** and then **sort by year**



2. Generate the variable size (ln of total asset is a good proxy for the size of the firm).

   Solution: you can generate the variable size by using the function **LN**

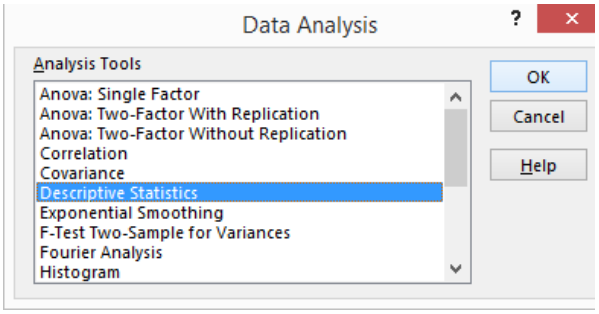
   Type "=ln(H2)" in the new column (called size) and then drag the column. Please see column I in the excel file "Article 6.1 – Solutions".

3. Some literature argues that if firms hold an amount of cash and cash equivalents not smaller than the 25% of their total asset, then they should be considered cash conservative. Construct a dummy variable, which takes the value of 1 if cash_to_asset is equal or greater than 0.25 and 0 otherwise.

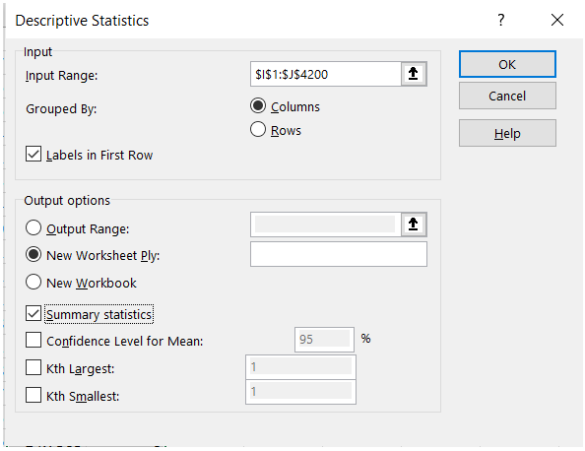
   Solution: you can generate the variable cash_conservative by using the function **IF**
   Type "=IF(G2>=0.25,1,0)" in the new column (called cash_con) and then drag the column. Please see column J in the excel file "Article 6.1 – Solutions".

4. Calculate the summary statistics for the two new variables: size and cash_con. Comment the values obtained. Which is the probability of being cash conservative firms?

   Solution: You need to turn on the statistical analysis add-in. Click **File** and then **Options ->** Click **Add-Ins** and select **Analysis ToolPak** (then click the GO button) -> Click on the **Data** tab and then double click the newly introduced button **Data Analysis ->** Choose **Descriptive Statistics**



Then:



Please see the descriptive statistics table in the sheet called “Descriptive Statistics” in the excel file “Article 6.1 – Solutions”.

The mean associated to the generated dummy (cash_con) shows the percentage of cash conservative firms in the sample. In our case, 11.31% of firms are cash conservative.

Otherwise you can calculate the summary statistics by using statistical functions:

Mean can be calculate using function AVERAGE

Median can be calculate using function MEDIAN

Mode can be calculate using function MODE

Variance can be calculate using function VAR.S

Skewness can be calculate using function SKEW

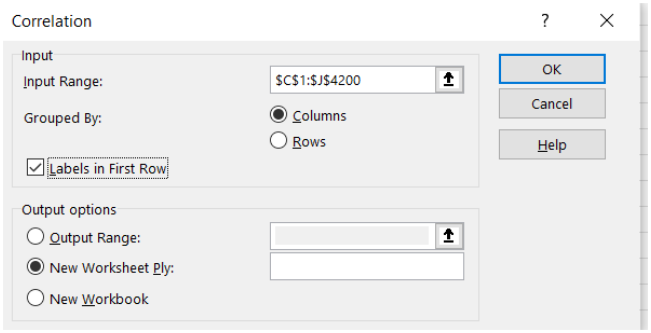
Kurtosis can be calculate using function KURT

Minimum can be calculate using function MIN

Maximum can be calculate using function MAX

5. Derive the correlation matrix for the main variables.

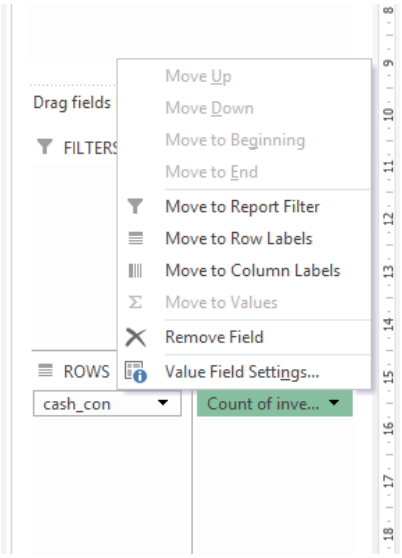
   Solution: Click on the **Data** tab and then double click the newly introduced button **Data Analysis ->** Choose **Correlation**



   Please see the correlation matrix in the sheet called "Correlation" in the excel file "Article 6.1 – Solutions".

6. Using the pivot table, calculate the average of investments for cash conservative and non-cash conservative firms.

   Solution: Select all our data -> click **Insert** tab and then **Pivot Table** -> select where you want the Pivot Table to be placed -> ok
   Select cash_con in rows and investment in values
   Click on investment (count of) and then on **Value Field Setting** -> select **Average**



   Please see the pivot table in the sheet called "Pivot Table 1" in the excel file "Data 3 – Solutions".

7. Create a dummy variable which takes the value of 1 if size is equal or greater than its average (big size firms) and 0 otherwise (small size firms)

Solution: you can generate the variable big_small by using the functions **IF** with **AVERAGE** (nested function).

Type "=IF(I2>=AVERAGE($I$2:$I$4200),1,0)" in the new column K (called big_small) and then drag the column. Please see column L in the excel file "Article 6.1 – Solutions".

8. Using the pivot table, calculate the average of investments for different sizes and for cash and non-cash conservative firms.

   Solution: Select all our data -> click **Insert** tab and then **Pivot Table** -> select where you want the Pivot Table to be placed -> ok
   Select big_small and cash_con in rows and investment in values
   Click on investment (count of) and then on **Value Field Setting** -> select **Average**

   Please see the pivot table in the sheet called "Pivot Table 2" in the excel file "Article 6.1 – Solutions".

9. Calculate the mean of the variable cashflow for each year. Plot the obtained values in an appropriate graph.

   Solution: Select all our data -> click **Insert** tab and then **Pivot Table** -> select where you want the Pivot Table to be placed -> ok
   Select year in rows and cashflow in values
   Click on investment (count of) and then on **Value Field Setting** -> select **Average**

   To make the appropriate graph:

   Click **Pivot Chart** in the **Analyze / Pivottable Tools** Tab -> select the most appropriate graph -> ok

   Please see the pivot table and the chart in the sheet called "Pivot Chart" in the excel file "Article 6.1 – Solutions".

10. Sheet2 contains the variable "Sales" which is the ln of net sales, and it is employed as a proxy of firms' revenues. Incorporate this variable into the main dataset (sheet "data") using VLOOKUP function.

    Solution: The VLOOKUP function uses the following arguments:

    - Lookup_value (required argument) – Lookup_value specifies the value that we want to look up in the first column of a table. VLOOKUP works in a left to right order, so you need to ensure that the information you want to look up is to the left of the corresponding data you want to extract. In this case, in order to combine two files, make sure you have a unique lookup value.
    - Table_array (required argument) – The table array is the data array that is to be searched. The VLOOKUP function searches in the left-most column of this array.

- Col_index_num (required argument) – This is an integer, specifying the column number of the supplied table_array, that you want to return a value from.
- Range_lookup (optional argument)

Create the lookupvalue by combining ID and YEAR

= A2&B2. You can also use the function CONCATENATE.

And then you dragged down for the entire column. You have to create this variable in both sheets (Data and Sheet2).

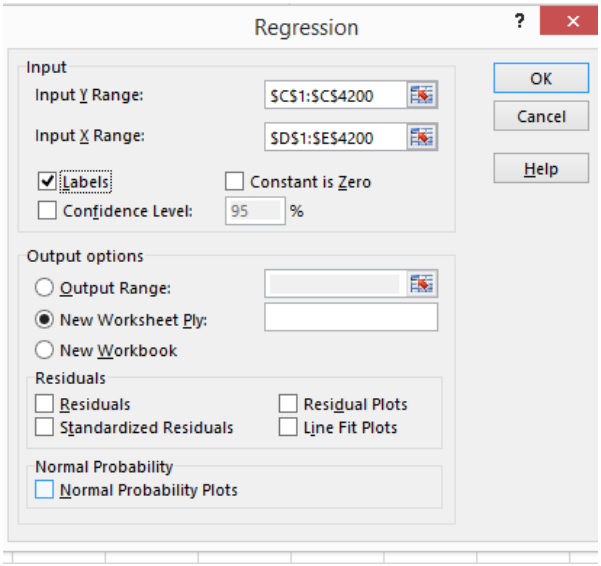
Then we can do the VLOOKUP. (to avoid potential error, you can also try to copy and past values for this new variable):

=VLOOKUP($L$2:$L$4200,'Sheet2 (2)'!$C$2:$D$4200,2,)

Please see columns L and M in the excel file "Article 6.1 – Solutions".".

11. Estimate the above equation using OLS:

Solution: Click on the **Data** tab and then double click the newly introduced button **Data Analysis ->** Choose Regression -> then select as follows:



Please see the regression in the sheet called "Regression" in the excel file "Article 6.1 – Solutions".